Can correcting feature location in simulated mean climate improve agreement on projected changes?

Adam A. L. Levy,[1] William Ingram,[1,2] Mark Jenkinson,[3] Chris Huntingford,[4] F. Hugo Lambert,[5] and Myles Allen[1,6]

Received 12 October 2012; revised 23 November 2012; accepted 7 December 2012; published 17 January 2013.

[1] To the extent that deficiencies in GCM simulations of precipitation are due to persistent errors of location and timing, correcting the spatial and seasonal distribution of features would provide a physically based improvement in inter-model agreement on future changes. We use a tool for the analysis of medical images to warp the precipitation climatologies of 14 General Circulation Models (GCMs) closer to a reanalysis of observations, rather than adjusting intensities locally as in conventional bias correction techniques. These warps are then applied to the same GCMs' simulated changes in mean climate under a CO_2 quadrupling experiment. We find that the warping process not only makes GCMs' historical climatologies more closely resemble reanalysis but also reduces the disagreement between the models' response to this external forcing. Developing a tool that is tailored for the specific requirements of climate fields may provide further improvement, particularly in combination with local bias correction techniques. **Citation:** Levy, A. A. L., W. Ingram, M. Jenkinson, C. Huntingford, F. H. Lambert, and M. Allen (2013), Can correcting feature location in simulated mean climate improve agreement on projected changes?, *Geophys. Res. Lett.*, *40*, 354–358, doi:10.1029/2012GL053964.

1. Introduction

[2] Precipitation is among the most important climate variables, and so understanding how it varies in response to climate change is a priority in projections for the future. However, the relatively small spatial scales of precipitation, combined with the complexities of the hydrological cycle, mean that even state of the art General Circulation Models (GCMs) not only struggle to agree on changes in precipitation distribution but also have difficulties recreating the current climate [*Dai*, 2006]. This has led to challenges for the detection and attribution of historical changes in precipitation [*Zhang et al*., 2007], as well as the prediction of trends under climate change scenarios [*Hegerl et al*., 2007]. Such biases can dramatically impact the mean climate of hydrological variables, as well as their projected changes, and so correction of such errors is vital for reliable hydrological impact assessment [*Ashfaq et al*., 2010].

[3] GCMs' projections will differ, partly due to differences in their physical responses to forcings but partly due to biases in the simulation of their mean climate. The former reflect model uncertainty as to the physical nature of the response but the latter do not. Such errors can be conceptually classed as spurious or missing features, wrong location of features [Allen and Ingram, 2002], and wrong intensity distribution of features, although typically no one of these dominates. Whilst improvements in GCM physics are needed to correct completely spurious features, biases within features can be corrected using bias correction techniques. Local bias correction methods are a powerful way of correcting precipitation intensity distributions by adjusting historical distributions to match observations [*Piani et al*., 2010; *Li et al*., 2010] but are less appropriate when the error is one of location [*Haerter et al*., 2011]. However, if errors in GCMs' spatial and seasonal distribution can be corrected, based on each GCM's simulated historical climate, subsequent local bias corrections may be smaller and have more physical basis in their use.

[4] While correcting locations of climatological features necessarily brings the GCMs' present-day climates closer together, it remains an interesting question whether it reduces the spread of their projected future changes. For example, most GCMs might simulate a feature's *response* to climate change in the same way but would appear to disagree if they all contain different errors in the mean location of the feature. Hence, correcting location biases can clarify how far differences between GCMs' projections are due to different physical responses, as opposed to different simulations of the mean climate.

[5] Although several spatial mapping tools have been developed for investigating precipitation distributions, these have only been used as weather forecast verification tools [*Brown et al*., 2012] and not for climate scales. We propose a new technique for the correction of projections, based upon the errors in the GCMs' mean historical climate. We make use of a publicly available (http://fsl.fmrib.ox.ac.uk/fsl/) medical image registration tool, *FNIRT* (FMRIB's Non-linear Image Registration Tool, part of the FMRIB Software Library) [*Woolrich et al*., 2009; *Andersson et al*., 2010], which was developed for the transformation of brain images taken from MRI (Magnetic Resonance Imaging) scans onto a common grid, so that inter-patient comparisons can be made.

[6] Using *FNIRT*, we derive transforms (or "warps") for 14 CMIP5 (Coupled Model Intercomparison Project Phase 5 (http://cmip-pcmdi.llnl.gov/)) GCMs (Table 1), which shift each model's precipitation climatology to more closely match

[1]Department of Physics, University of Oxford, Oxford, UK.
[2]Met Office Hadley Centre, Exeter, UK.
[3]Oxford Centre for Functional MRI of the Brain, University of Oxford, Oxford, UK.
[4]Centre for Ecology & Hydrology, Wallingford, UK.
[5]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.
[6]School of Geography and the Environment, University of Oxford, Oxford, UK.

Corresponding author: A. A. L. Levy, AOPP, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. (levy@atm.ox.ac.uk)

0094-8276/13/2012GL053964

Table 1. List of Models Used in Study

Model Name(s)	Modeling Group
CNRM-CM5	Centre National de Recherches Meteorologiques
CanESM2	Canadian Center for Climate Modelling and Analysis
GFDL-ESM2G, GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory
HadGEM2-ES	Met Office Hadley Centre
INM-CM4	Institute for Numerical Mathematics
IPSL-CM5A-LR, IPSL-CM5A-MR	Institut Pierre-Simon Laplace
MIROC-ESM, MIROC5	Japan Agency for Marine-Earth Science and Technology
MPI-ESM-LR, MPI-ESM-P	Max Planck Institute for Meteorology
MRI-CGCM3	Meteorological Research Institute
NorESM1-M	Norwegian Climate Center

the observed distribution. These warps are then applied to each GCM's projected changes under a prescribed external forcing, to evaluate whether greater agreement is reached through this technique than with uncorrected model outputs.

2. Methods

2.1. Medical Image Registration

[7] Medical image registration operates by warping an input image to better resemble a reference image. For inter-patient registration, this requires a non-linear, reversible (diffeomorphic) transform [*Beg et al.*, 2005] that can distort an image smoothly but cannot create folds or tears. One approach to determine this warp is to represent it with basis functions [*Rueckert et al.*, 1999; *Ashburner and Friston*, 1999] and find the coefficients (parameters) by numerically minimizing an objective function of the form [*Bajcsy and Kovačič*, 1989]

$$\sum_i [C(W_i(I), R_i) + \lambda O(W_i)] \quad (1)$$

where the i indexes grid points, I and R are the input and reference images, respectively, W is the warp, C is a cost function, which quantifies the difference between the reference image and the warped input image, λ is a tuneable parameter, and O is a regularization function, penalizing warps that are not spatially smooth [*Andersson et al.*, 2010].

[8] In *FNIRT*, we set the cost function, C, to take the form of a sum of squared differences, and the regularization term, O, is the mechanical bending energy of an elastic solid subject to the warp W [*Ashburner and Friston*, 1999]. The tuning of λ determines how smooth the warp will be, as it determines how great a role O plays in the objective function.

2.2. Applying FNIRT to Climate Models

[9] *FNIRT* conserves intensities (in our case, precipitation rates) when warping, and so total rainfall is *not* conserved. In addition, FNIRT is designed to warp medical images of patients' brains and so does not have the periodic boundary conditions that our problem requires both longitudinally and seasonally, in order to allow features to move across the prime meridian as well as between late and early seasons (e.g., between December and January). *FNIRT* naturally operates on a three-dimensional Cartesian geometry with no periodicity. For a preliminary investigation of the potential of these methods, we therefore align the longitudinal, latitudinal, and temporal dimensions along the rectilinear x, y, and z

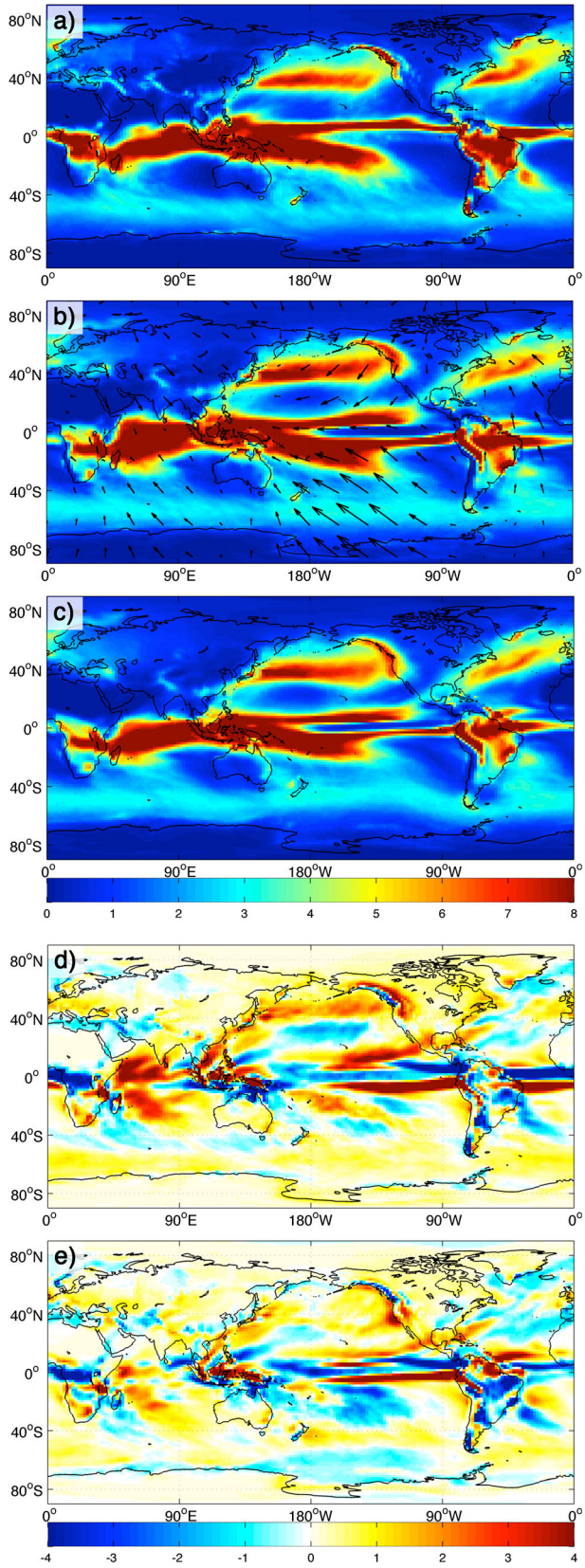

Figure 1. Example of the warping process, showing precipitation in millimeters per day for January and one GCM, INM-CM4. Panels illustrate (a) ERA-Interim, (b) INM-CM4 with warp vectors, (c) INM-CM4 after warping, (d) INM-CM4 (before warping) minus ERA-Interim, and (e) INM-CM4 (after warping) minus ERA-Interim.

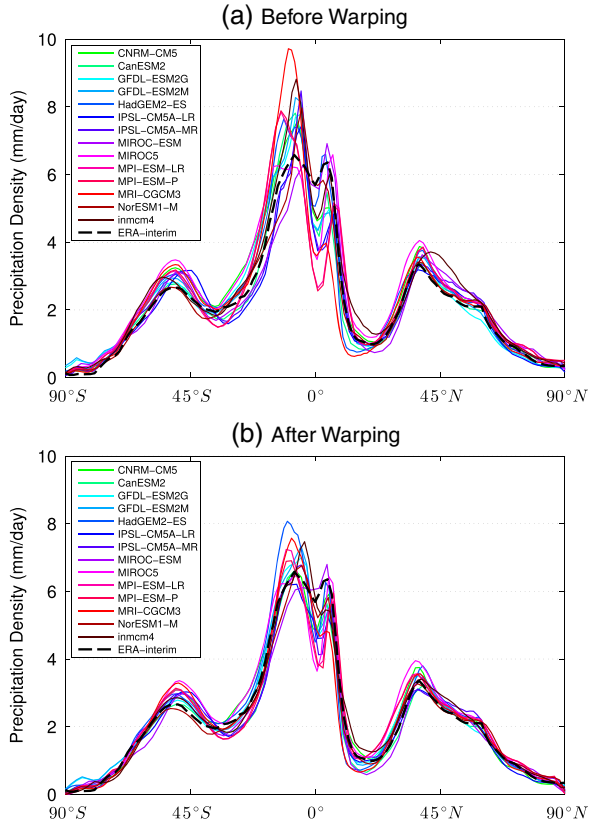

Figure 2. Historical zonal mean for January for all models used (see legend) with latitude on the *x* axis, before (a) and after (b) the warping process is applied.

dimensions, respectively, padding each climatology with copies of itself at the longitudinal and temporal boundaries. This allows the central warp to move features across the boundaries, approximating the behavior of true periodic boundary conditions. As there is little precipitation at high latitudes, we do not address the polar boundary condition. In our grid, we define 1 month as equivalent to 10° of latitude or longitude, chosen as giving the greatest improvement (not shown) to the climatology of HadCM3 (a model from the previous generation of GCMs).

[10] For our reference, R in equation (1), we take the monthly means of the ERA-Interim reanalysis [*Dee et al.*, 2011] of January 1979 to December 2005 interpolated to each GCM's grid. Although reanalysis datasets do not precisely match observations, they are much closer to reality than GCMs, and their complete coverage is convenient for our purposes. We take each model's monthly historical climatology for the same period as our input, I in (1), as sampling uncertainty in such 27 year means will be small compared to the GCMs' systematic errors. We repeat the warping process iteratively, until convergence is reached, for a range of different values of λ, varying from 25 to 200 $(\text{mm d}^{-1})^2$. The warps derived using a λ of 150 $(\text{mm d}^{-1})^2$ were selected, as (on average) they maximized the correlation between the models and ERA-Interim's climatology, whilst preventing features from entering on both the Eastern and Western boundaries and avoiding regions where the warp's Jacobian becomes very large or small (the Jacobian of a transformation quantifies the extent to which regions are stretched and compressed).

[11] Figure 1 illustrates how 1 month (January) of one model (INM-CM4) is modified under the registration process. One conspicuous error in this model (Figure 1d) is the "split ITCZ"—a common error in GCMs [*Dai*, 2006]. However, as the warp can stretch and distort, but cannot cause regions to vanish, the warp is limited to shifting the lower branch West, as well as compressing it latitudinally (Figure 1b). The high value of λ imposes a smoothness on the warp, which restricts its ability to shift the Northern ITCZ East. The level of smoothness required can lead to increased discrepancies with reanalysis in some areas, as seen in the North Atlantic storm tracks. Another conspicuous error, the over-extensive region of heavy rain over the Indian Ocean shown in Figure 1b, is more amenable to warping and is shrunk down to closely resemble the ERA-Interim pattern (as illustrated by the reduction in red for this region in Figure 1d relative to 1e).

[12] To investigate the potential of this technique to remove avoidable uncertainty from projections, we now consider the change in precipitation climatology predicted by the models under a prescribed external forcing. Errors of location will not remain totally unchanged through climate change, but our hope is that any substantial error in the present-day simulation will be physically persistent enough that such a correction will continue to remove more error than it introduces. We choose the CMIP5 1% CO_2 experiment, in which carbon dioxide levels are increased by 1% each year until they have reached quadruple their initial values (140 years), as such a strong forcing will have a high signal-to-noise ratio. For each GCM, we evaluate the monthly mean climatology for the first and last 30 years of the experiment (a separation time of 110 years) and take the difference. We then apply the GCMs'

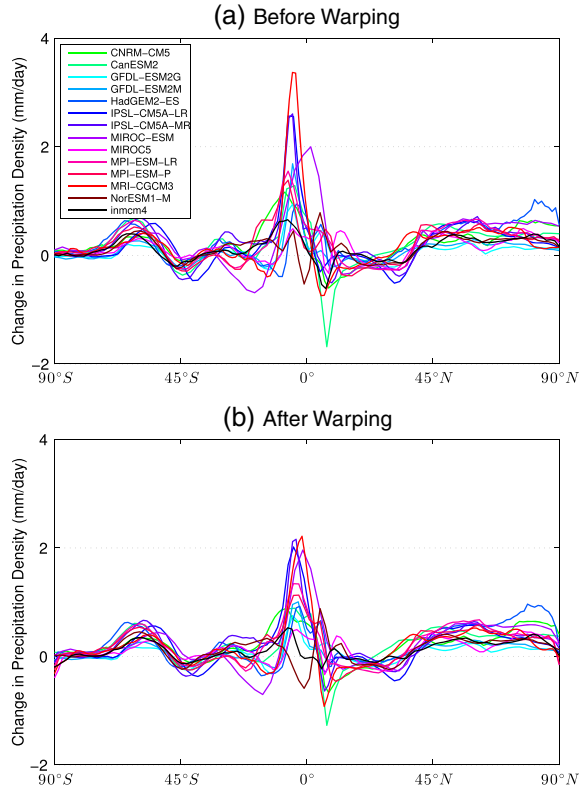

Figure 3. Zonal mean change in precipitation for January for all models used (see legend) under the 1% CO_2 experiment, with latitude on the *x* axis, before (a) and after (b) warping. All models have a reduced root mean square difference from the multi-model mean.

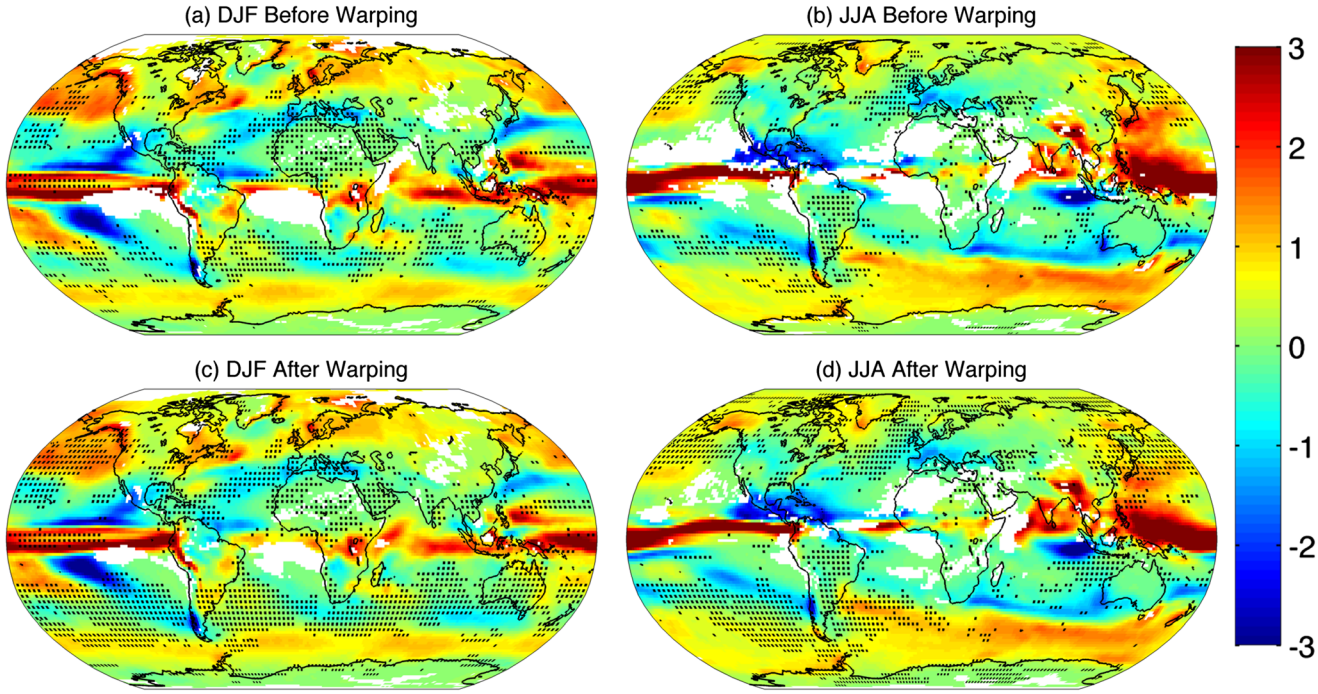

Figure 4. Projected change in precipitation seasonal climatology over 110 years under the 1% CO_2 CMIP5 experiment (mean of first 30 years, subtracted from the mean of last 30 years), displayed in millimeters per day. Color indicates an inter-model standard deviation smaller than the historical inter-annual variability at that location, and stippling denotes a standard deviation smaller than the estimated inter-decadal variability. An increase in color and stippling therefore illustrate greater agreement between GCMs. The panels show December–February (a and c) and June–August (b and d), both before (a and b) and after (c and d) warping.

warps, derived purely from the historical climate, and investigate how the agreement between models changes as a result.

3. Results

[13] Warping the models' mean monthly climatologies to ERA-Interim with a λ of 150 $(\text{mm day}^{-1})^2$ reduces the error (difference from 1) in their correlation with the reanalysis by 35%, on average. Figure 2 shows this improvement and verifies that applying our warping technique to GCMs *does* shift their mean climatologies closer to the observations, as it should by construction.

[14] We next evaluate the variation within the GCMs' projected changes under the 1% CO_2 scenario. We quantify inter-model agreement using the standard deviation and range at each point on the globe, after the models have been resampled to a 1° grid for utmost detail, although spatial smoothing can provide more information in the sense of increasing the signal to noise (C. F. McSweeney and R. G. Jones, No consensus on consensus: The challenge of finding a universal approach to measuring and mapping ensemble consistency in GCM projections, submitted to *Climatic Change*, 2012). We find that the standard deviation and range are both reduced by 15%, after we apply the historically derived warps, which is exemplified by the zonal means shown in Figure 3. This improvement covers much of the globe, with 71% of the world seeing a reduction in standard deviation, and is also substantial in some regions, more than halving standard deviation for 5% of the Earth's surface.

[15] In contrast, a crude local bias correction (scaling the mean precipitation of each month to match the local value of the reanalysis—the only correction possible if one has only mean fields) increases inter-model standard deviation unless the scaling is limited to a very restricted range (a factor of 1.5 or less), in which case it makes little difference, with the global mean reduction never exceeding 2.6%. More sophisticated local techniques might perform better but are less appropriate if a feature, such as a convergence zone, is in the wrong place, rather than of the wrong magnitude [*Haerter et al.*, 2011], as this may introduce spuriously large intensity changes. However, we could physically expect judicious combination of the two classes of correction to out-perform either alone.

[16] *Hegerl et al.* [2007, Figure 10.9] use a single plot to represent the changes projected by the models, as well as a measure of the inter-model agreement, which is evaluated based on the *sign* of changes across GCMs. *Tebaldi et al.* [2011] argue that quantifying the inter-model agreement relative to an absolute value fails to allow for the range in natural variation across the globe and that comparing it to the signal itself can be misleading when the signal is small, whereas C. F. McSweeney and R. G. Jones (No consensus on consensus: The challenge of finding a universal approach to measuring and mapping ensemble consistency in GCM projections, submitted to *Climatic Change*, 2012) additionally stress the importance of such figures' legibility. We therefore define the agreement thresholds at each geographical location with respect to the historical inter-annual and estimated inter-decadal variability. These thresholds are chosen as the size of variation in the local climate is intuitive to those living in that region [*Tebaldi et al.*, 2011]. These thresholds are taken to be ERA-Interim's inter-annual standard deviation and $1/\sqrt{10}$ of this value, respectively. This is equivalent to criterion (1) of *Tebaldi et al.* [2011]. We use this technique to represent both summer and winter, before and after the warp is applied (Figure 4).

[17] Figure 4 demonstrates that although the distribution of precipitation changes averaged across models is not greatly altered by the warping technique, the technique does improve inter-model agreement (as measured by standard deviation). This can be seen by the increase in color and especially stippling in several regions in Figures 4c and 4d, relative to Figures 4a and 4b. There are some regions where agreement deteriorates but such regions are considerably smaller than the regions where agreement increases.

4. Conclusions

[18] We have demonstrated that techniques used in medical image registration can be adapted and applied to this problem in climate science. Warping GCMs to reanalysis of the observations not only improves their ability to recreate the historical precipitation distribution but also creates greater inter-model agreement on projected changes under a prescribed forcing (the CMIP5 1% CO_2 experiment). The latter result confirms that GCM errors of location do tend to persist under simulated climate change and that such errors can mask agreement within GCMs' responses to forcings. In contrast, a simple local bias correction technique increased inter-model discrepancies, unless so restricted that it had little effect, illustrating the difficulties in applying such techniques when the errors are location based.

[19] Improvements are most notable over sea, particularly the North Pacific, Southern subtropical oceans and Southern Ocean, although Australia in summer and East Asia in winter also show improvements (Figure 4). These areas of improvement result from the relatively large value for λ that was required for physical and numerical constraints and constrains warps to be smooth over larger scales. This limits warps to shifting larger features, such as the ITCZ, and prevents great improvements to smaller scale features, such as those of orographic precipitation.

[20] *FNIRT*'s inability to deal with the spherical geometry of this problem not only limits our ability to deal with periodic boundary conditions but also means that the areas of features could not be conserved when they are moved latitudinally (all $1^\circ \times 1^\circ$ grid squares are treated as equal area). Further, it is not clear whether it is more meaningful to conserve intensities or total precipitation upon warping, but *FNIRT* currently limits our investigation to conservation of precipitation density.

[21] We are therefore developing a registration tool designed specifically for this climatological problem, which we plan to make freely available. Its features will include operation in the correct geometry of our problem (handling the spherical geometry of the Earth's surface, as well as the seasonal cyclicity), the option to conserve total precipitation instead of precipitation rate, and the ability to weight regions (such as all land or a region of specific interest such as the Sahel), so they can be prioritized by the warp. These changes will also allow us to find less smooth transformations, and it is our hope that this will allow for closer mapping to observations, particularly for features with smaller scales. We also intend to allow for the derivation of warps based on a multivariate field, which would allow us to maintain physical consistency between a set of variables.

[22] The revised method will be applied to a range of more realistic climate forcing experiments, such as the IPCC's RCP scenarios, and could be combined with local bias corrections, to correct errors in both the intensity and spatial distributions of features. As well as achieving more physically and statistically robust estimates of future changes through such applications, we aim to apply this technique to the detection and attribution of climate change and hope that such applications will clarify our understanding of precipitation projections and so inform future policy and adaptation.

[23] **Acknowledgments.** We acknowledge the World Climate Research Program's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (Table 1) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and leads development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We would also like to thank Gill Martin and Rob Chadwick for invaluable advice on HadGEM2. AALL, WJI, MJ, CH, and FHL were supported by NERC under contract NE/I00680X/1 (HYDRA). WJI also received support from the Joint DECC/Defra Met Office Hadley Centre Climate Program (GA01101), and MRA from the NOAA/DoE IDAG project.

References

Allen, M. R., and W. J. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, *419*(6903), 224–232.

Andersson, J., M. Jenkinson, and S. Smith (2010), Non-linear registration, aka spatial normalisation. FMRIB technical report, available from http://fsl.fmrib.ox.ac.uk/analysis/techrep/tr07ja1/

Ashburner, J., and K. J. Friston (1999), Nonlinear spatial normalization using basis functions, *Hum. Brain Mapp.*, *7*, 254–266.

Ashfaq, M., L. C. Bowling, K. Cherkauer, J. S. Pal, and N. S. Diffenbaugh (2010), Influence of climate model biases and daily-scale temperature and precipitation events on hydrological impacts assessment: A case study of the United States, *J. Geophys. Res.*, *115*(D14), D14,116, doi:10.1029/2009JD012965.

Bajcsy, R., and S. Kovačič (1989), Multiresolution elastic matching, *Comput Vis Graph Image Process*, *46*(1), 1–21.

Beg, M. F., M. I. Miller, A. Trouvé, and L. Younes (2005), Computing large deformation metric mappings via geodesic flows of diffeomorphisms, *Int J Comput Vis.*, *61*(2), 139–157, doi:10.1023/B:VISI.0000043755.93987.aa.

Brown, B. G., E. Gilleland, and E. E. Ebert (2012), Forecast of spatial fields, in *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, second ed., chap. 6, John Wiley & Sons, Ltd.

Dai, A. (2006), Precipitation characteristics in eighteen coupled climate models, *J. Climate*, *19*(18), 4605–4630, doi:10.1175/JCLI3884.1.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, and F. Vitart (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q J Roy Meteorol Soc.*, *137*(656), 553–597.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani (2011), Climate model bias correction and the role of timescales, *Hydrol Earth Syst Sci*, *15*(3), 1065–1079, doi:10.5194/hess-15-1065-2011.

Hegerl, G. C., F. W. Zwiers, P. Braconnot, N. P. Gillett, Y. Luo, J. A. M. Orsini, N. Nicholls, J. E. Penner, and P. A. Stott (2007), Understanding and attributing climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge University Press.

Li, H., J. Sheffield, and E. F. Wood (2010), Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, *J. Geophys. Res.*, *115*(D10), D10,101, doi:10.1029/2009JD012882.

Piani, C., J. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, *99*, 187–192, 10.1007/s00704-009-0134-9.

Rueckert, D., L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes (1999), Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging*, *18*(8), 712–721.

Tebaldi, C., J. M. Arblaster, and R. Knutti (2011), Mapping model agreement on future climate projections, *Geophys. Res. Lett.*, *38*(23), L23,701, doi:10.1029/2011GL049863.

Woolrich, M., S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, C. B. T. Behrens, M. Jenkinson, and S. Smith (2009), Bayesian analysis of neuroimaging data in FSL, *NeuroImage*, *45*(1, Supplement 1), S173–S186.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa (2007), Detection of human influence on twentieth-century precipitation trends, *Nature*, *448*(7152), 461–465.